# BAYESIAN CLASSIFICATION OF SHAPES HIDDEN IN POINT CLOUD DATA

*Anuj Srivastava*[1] *and Ian H. Jermyn*[2]

[1]Department of Statistics, Florida State University, Tallahassee, FL, USA
[2]Project ARIANA, INRIA, Sophia Antipolis, France

## ABSTRACT

An interesting challenge in image processing is to classify shapes of polygons formed by selecting and ordering points in a 2D cluttered point cloud. This kind of data can result, for example, from a simple preprocessing of images containing objects with prominent boundaries. Taking an analysis-by-synthesis approach, we simulate high-probability configurations of sampled contours using models learnt from the training data to evaluate the given test data. To facilitate simulations, we develop statistical models for sources of (nuisance) variability: (i) shape variations of contours within classes, (ii) variability in sampling continuous curves into points, (iii) pose and scale variability, (iv) observation noise, and (v) points introduced by clutter. Finally, using a Monte Carlo approach, we estimate the posterior probabilities of different classes which leads to a Bayesian classification.

***Index Terms*—** shape models, Bayesian shape estimation, clutter model, Monte Carlo inference

## 1. INTRODUCTION

The classification and recognition of objects in images is an important problem in biometrics, medical image analysis, and many other branches of science. A common approach is to represent the objects of interest with certain discriminant features, and then use some statistical models on these feature spaces for classification. An important feature of many objects is their **shape** and, as a consequence, shape analysis has become an integral part of object classification [1]. One way to use shape analysis is to estimate the boundaries of the objects (in images) and to analyze the shapes of those boundaries in order to characterize the original objects. Towards that end, there have been several papers in the literature on analyzing the shapes of continuous, closed, planar curves (see for example [2, 3] and others referenced therein). While such continuous formulations are fundamental in understanding shapes and their variability, practical situations mostly involve heavily under-sampled, noisy, and cluttered discrete data, often because the process of estimating boundaries uses low-level techniques that extract a set of primitives (points, edges, arcs, etc.) in the image plane. We will restrict attention to points in this paper, but the method generalizes to more complex primitives. Therefore, an important problem in object recognition is to relate (probabilistically) a given set of primitives to pre-determined (continuous) shape classes and to classify this set using a *fully statistical framework*.

The research presented here was partially supported by ARO W911NF-04-01-0268, AFOSR FA9550-06-1-0324, and Northrop-Grumman Innovation Alliance grants, and by the INRIA/FSU Associated Team "SHAPES" grant.

### 1.1. Problem Challenges

The biggest challenge is to select and organize a large subset of the given primitives into shapes that resemble the shapes of interest. Through an example presented in Figure 1, we will explain these components. The number of permutations for organizing primitives into shapes is huge. For example, if we take the primitives to be points, the number of possible polygons using $40$ distinct points is of the order of $10^{47}$. If we select only $20$ points out of the given $40$ and form a polygonal shape, the number of possibilities is still $10^{29}$. To form and evaluate all these shape permutations is impossible. Our solution is to analyze these configurations through synthesis, *i.e. to synthesize high-probability configurations from known shape classes and then to measure their similarities with the data*. Although this approach has far smaller complexity than the bottom-up combinatoric approach, the joint variability of all the unknowns is still enormous. To go further, one must use the structure of the problem to break down the variability into components, and then probabilistically model the components individually. Through an example presented in Figure 1, we will try to explain these components.

## 2. PROBLEM FORMULATION AND OVERVIEW

The classification problem is described by the probability $P(C|\mathbf{y})$, where $C \in \mathcal{C}$ is the class of the object represented by the data set, and $\mathbf{y} \subset \mathcal{Y}$ is the data, *i.e.* a finite set of primitives. (Because we are restricting attention to primitives that are simply points in $\mathbb{R}^2$, we have $\mathcal{Y} = \mathbb{R}^{2m}$ for $m$ primitives.) Classification can then be performed by maximizing the probability: $\hat{C} = \operatorname{argmax}_C P(C|\mathbf{y})$. The difficulty of the problem is contained in $P(\mathbf{y}|C)$, which describes the formation of the data starting from the object class. To make

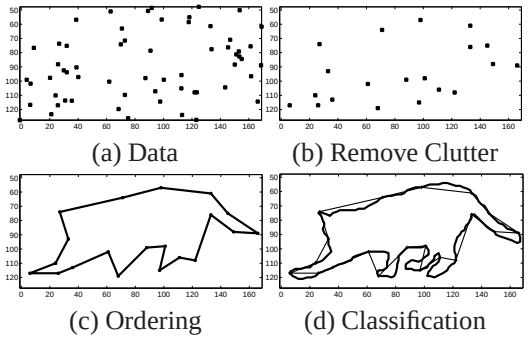



**Fig. 1**. Problem Challenges: The point cloud in (a) contains clutter as well as the shape of interest. The removal of clutter leads to the points in (b), which when ordered result in a polygon (c). Subsequently, this polygon can be used for shape classification, as in (d).

any further progress, this probability must be broken down into components corresponding to simpler stages in the data formation process.

First, we introduce some variables: Let $g \in \mathcal{G}$, where $\mathcal{G} \equiv (SO(2) \ltimes \mathbb{R}^2) \times \mathbb{R}_+$, be a similarity transformation that includes rotation, translation, and scale. The symbol $\ltimes$ denotes the semi-direct product. Let $q \in \mathcal{Q}$ be a shape, *i.e.* an object boundary modulo similarity transformations and reparameterizations. Thus, a specific boundary is given by $gq$. Let $s \in \mathcal{S}$ represent $n$ point-primitives on the shape boundary; among other variables $s$ contains $n$. We will call this a "sampling". Then $qs$ will be a set of $n$ point primitives modulo a similarity transformation, while a specific set of point primitives is given by $\mathbf{x} = gqs$. Finally, let $\mathcal{I} \ni \iota : [0, \ldots n] \rightarrow [0, \ldots m]$ be a one-to-one map, *i.e.* an injection, relating each element of $\mathbf{x}$ to a unique element of $\mathbf{y}$.

Now, we can write (making certain independence assumptions, to be discussed later)

$$P(\mathbf{y}|C) = \sum_{\iota \in \mathcal{I}} \iiint_{\substack{g \in \mathcal{G} \\ s \in \mathcal{S} \\ q \in \mathcal{Q}}} P(\mathbf{y}|\iota, gqs)\; P(\iota|s)\; P(g|q, C) \times P(s|q, C)\; P(q|C) dg\; ds\; dq\; . \quad (1)$$

We will take $P(\iota|s)$ and $P(g|q, C)$ to be uniform. With these assumptions, $g$ and $\iota$ appear solely in the first factor in the integrand, $P(\mathbf{y}|\iota, gqs)$.

Our algorithmic strategy for dealing with this complexity is based on two approximate methods for evaluating the integrals and sums: Monte Carlo integration and the Laplace's method. We use the first for the integrals over $q$ and $s$, generating realizations from their probability distributions and then summing the values of the integrand evaluated at these realizations. We use the second for the integral over $g$ and the sum over $\iota$. Using a combination of the Hungarian algorithm and Procrustes alignment, we solve the joint registration ($g$)-transformation ($injection$) problem. The cost function for this optimization is the likelihood $P(\mathbf{y}|\iota, gqs)$, which must include a stochastic model of the clutter points. The result of these procedures is an approximation to $P(\mathbf{y}|C)$ for each value of $C$, *i.e.* each class, and thus, after a trivial normalization, to the value of $P(C|\mathbf{y})$. Classification is then immediate.

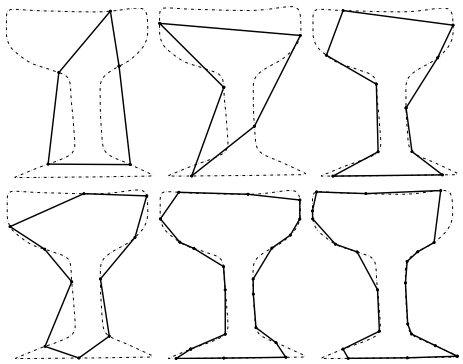


**Fig. 2**. Illustration of sampling variability for a curve.

To construct a fully statistical framework, then, we have to develop probability models and computational methods for the variability in shape ($P(q|C)$), sampling ($P(s|q, C)$), and observation noise and clutter ($P(\mathbf{y}|\iota, gqs)$). We now discuss each of these in more detail, beginning with sampling, since our approach here is novel.

## 3. MODELING SAMPLING VARIABILITY

By a sampling of a continuous curve, we mean selecting an ordered finite number of points on that curve. (We underline the distinction between our use of "sampling a continuous curve" and the phrase "sampling from a probability". To avoid this confusion, we will use "simulation" for the latter.) The sampling step results in a loss of information about the original shape. Figure 2 shows some examples of samplings of a single shape. Since the sampled points are ordered, we can draw a polygon to improve the visualization of the sampled points.

### 3.1. Representation

How can we mathematically represent a sampling? The process of sampling, by itself, is seldom studied in the literature, although the related problem of matching sampled shapes has received a lot of attention. A sampling involves two elements: a certain number of points, $n$, and their placement on the curve. The latter can be expressed by parameterizing the curve in terms of its arc length, and then selecting $n$ values in the interval $[0, L]$, where $L$ is the length of the curve. Since we will be sampling the points from shapes, we can assume

that $L = 1$. Note that this assumes that the probability of a sampling does not depend on the position, orientation, and scale of a curve, which is implicit in Eqn. 1.

Let $\Gamma$ be the set of increasing, differentiable functions from $[0,1]$ to itself, such that for all $\gamma \in \Gamma$, $\gamma(0) = 0$ and $\gamma(1) = 1$, or, in other words, the group $\Gamma$ of positive diffeomorphisms of the unit interval. Now let $U = [0 \ldots n]/n$ be a uniform partition of the interval $[0,1]$ into $n$ sub-intervals. A sampling $s$ will be represented by an equivalence class of triples $\langle n, \tau, \gamma \rangle \in \mathbb{N} \times \mathbb{S}^1 \times \Gamma$, with the actual samples on a (arc-length parameterized, unit length curve) $\beta$ being $\beta(\tau + \gamma(0))$, $\beta((\tau + \gamma(1/n))$,..., $\beta((\tau + \gamma(1))$. The advantage of this representation is that we can change $n$ without changing $\gamma$, and vice-versa. We still have to decide, however, how to represent $\gamma$. The functions in $\Gamma$ can be thought of as cumulative distribution functions for nowhere-zero probability densities on $[0,1]$, with which they are in bijective correspondence, and this gives rise to a number of possibilities for representing such functions:

**Diffeomorphism**: An element of $\Gamma$ is represented as itself, *i.e.* as an increasing function from $[0,1]$ to itself, such that $\gamma(0) = 0$ and $\gamma(1) = 1$. The advantage of this representation is that the action of the group of diffeomorphisms on itself is particularly simple, by composition.

**Probability density**: Here an element of $\Gamma$ is represented by its derivative, denoted $\mathcal{P} \ni p = \dot{\gamma}$, which is an everywhere positive probability density on $[0,1]$, *i.e.* a positive function that integrates to 1.

**Log probability**: Here an element of $\Gamma$ is represented by the logarithm of a probability density. It is an arbitrary function whose exponential integrates to 1. The advantage of this representation is that the values of log-probability function are unconstrained, apart from the overall normalization.

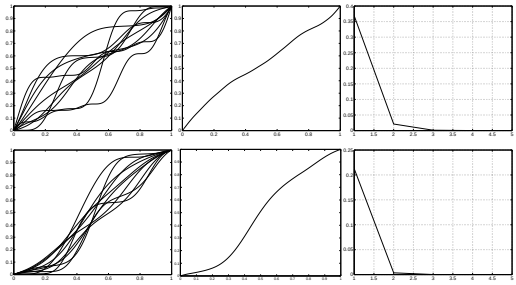
**Square-Root Form**: An element of $\Gamma$ is represented by the square root of a probability density, $\Psi \ni \psi = p^{\frac{1}{2}}$. This is a positive function whose square integrates to 1, *i.e.* its $\mathbb{L}^2$ norm is 1. The set of these functions thus forms the positive orthant of the unit sphere in the space $\mathbb{L}^2([0,1])$. The advantage of this representation is that it greatly simplifies the form of the most natural Riemannian metric one can place on $\Gamma$, as we will now discuss.

### 3.2. Riemannian Structure on $\Gamma$

While there are clearly a large number of Riemannian metrics one could place on $\Gamma$, it is a remarkable fact, proved by Čencov [4], that on spaces of probability distributions on finite sets, there is a unique Riemannian metric that is invariant to "Markov mappings". This Riemannian metric is the so called *Fisher-Rao* (F-R) metric. The F-R metric extends naturally to the space of probability measures on continuous spaces such as $[0,1]$, where it is invariant to the (reparameterization) action of the diffeomorphism group. Since $\Gamma$ is isomorphic to the space of probability measures, we can view the F-R metric as a metric on $\Gamma$ too. Because of its invariance properties, this is the metric we choose to use. In terms of the probability density representation, it takes the following form: the inner product between tangent vectors $\delta p$ and $\delta' p$ to the space of probability distributions on $[0,1]$ (here tangent vectors are functions that integrate to zero) at the point $p \in \mathcal{P}$ is $\langle \delta p, \delta' p \rangle_p = \int_0^1 \delta p(s)\ \delta' p(s)\ \frac{1}{p(s)}\ ds$. It turns out, however, that the F-R metric simplifies greatly under the half-density representation. Indeed, it becomes $\mathbb{L}^2$, because $\psi^2 = p$ means that $2\psi\delta\psi = \delta p$, and thus that $\langle \delta\psi, \delta'\psi \rangle_\psi = \int_0^1 \delta\psi(s)\ \delta'\psi(s)\ ds$. We have already seen that $\Psi$ is the positive orthant of the unit sphere in $\mathbb{L}^2([0,1])$, and now we see that the F-R metric is simply the $\mathbb{L}^2$ Riemannian metric on $\mathbb{L}^2([0,1])$ restricted to $\Psi$. The space $\Psi$ endowed with the F-R metric is thus the positive orthant of the unit sphere in $\mathbb{L}^2([0,1])$ with the induced Riemannian metric. Consequently, geodesics under the F-R metric are nothing but great circles on this sphere, while geodesic lengths are simply the lengths of shortest arcs on the sphere. Arc-length distance on a unit sphere has been used to measure divergences between probability density functions for a long time [5]. This metric also plays an important role in information geometry as developed by Amari [6].

Now we list some analytical expressions that are useful for statistical analysis on $\Psi$ and thus on $\Gamma$. As $\Psi$ is an infinite-dimensional sphere inside $\mathbb{L}^2([0,1])$, the length of the geodesic in $\Gamma$ between any two functions $\gamma_1$ and $\gamma_2$ under the F-R metric is given by $d(\gamma_1, \gamma_2) = \cos^{-1}(\left\langle \dot{\gamma}_1^{\frac{1}{2}}, \dot{\gamma}_2^{\frac{1}{2}} \right\rangle)$, where the inner product is $\mathbb{L}^2$. The geodesic between two points $\gamma_1$ and $\gamma_2$ of $\Gamma$ is similarly derived. For $\psi_i = \dot{\gamma}_i^{\frac{1}{2}}$, the corresponding geodesic in $\Psi$ is given by $\psi(t) = \frac{1}{\sin(\theta)}\left[\sin((1-t)\theta)\psi_1 + \sin(t\theta)\psi_2\right]$, where $\cos(\theta) = \langle \psi_1, \psi_2 \rangle$. The desired geodesic in $\Gamma$ is then given by $\gamma(t)$, where $\gamma(t)(s) = \int_0^s \psi(t)(\tau)^2\ d\tau$. Due to this additional integration step, it is sometimes easier to perform the Riemannian analysis in $\Psi$ and to map the final result back to $\Gamma$. This is especially true for computing means and variances of sampling functions, for constructing probability densities on $\Gamma$, and for simulating from these probability densities.

In $\Psi$, the geodesic starting from a point $\psi$, in the direction $v \in T_\psi(\Psi)$, can be written as: $\cos(t)\psi + \sin(t)\frac{v}{\|v\|}$ (with the $\mathbb{L}^2$ norm). As a result, the exponential map, $\exp : T_\psi(\Psi) \to \Psi$, has a very simple expression: $\exp_\psi(v) = \cos(\|v\|)\psi + \sin(\|v\|)\frac{v}{\|v\|}$. The exponential map is a bijection between a tangent space and the unit sphere if we restrict $\|v\|$ so that $\|v\| \in [0, \pi)$, but for large enough $\|v\|$, $\exp_\psi(v)$ will lie outside $\Psi$, *i.e.* $\psi$ may take on negative values. We will discuss this further when we define prior probabilities on $\Gamma$. For any $\psi_1, \psi_2 \in \Psi$, we define $v \in T_{\psi_1}(\Psi)$ to be the inverse exponential of $\psi_2$ if $\exp_{\psi_1}(v) = \psi_2$; we will use the notation $\exp_{\psi_1}^{-1}(\psi_2) = v$. This can be computed using the following steps: $u = \psi_2 - \langle \psi_2, \psi_1 \rangle \psi_1$, $v =$



(a) (b) (c)

**Fig. 3**. Examples of Karcher means in $\Gamma$: In each case, (a) shows ten $\gamma_i$, (b) shows their Karcher mean $\mu_\gamma$, and (c) shows the cost functions vs. iterations.

$u\cos^{-1}(\langle\psi_1,\psi_2\rangle)/\langle u,u\rangle^{\frac{1}{2}}$.

### 3.3. Statistics on $\Gamma$

Consider the task of computing the statistical mean of a set of sampling functions $\{\gamma_1,\gamma_2,\ldots,\gamma_k\}$ intrinsically in $\Gamma$. As mentioned earlier, we will use the square-root forms of these functions to perform such calculations. Let the corresponding set of square-root forms be given by $\{\psi_1,\psi_2,\ldots,\psi_k\}$, $\psi_i = \dot{\gamma}_i^{\frac{1}{2}}$. We define their Karcher mean as:

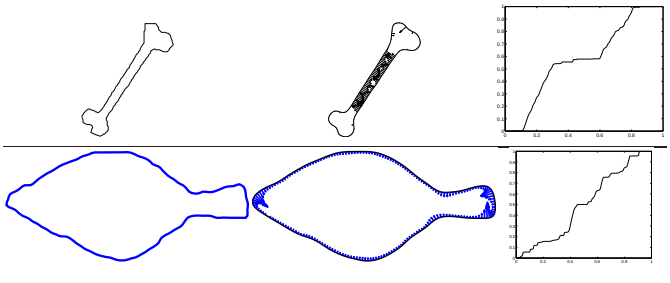
$$\mu = \underset{\psi\in\Psi}{\operatorname{argmin}} \sum_{i=1}^{k} d(\psi,\psi_i)^2 \ ,$$

where $d$ is the geodesic distance on $\Psi$. The minimum value $\sum_{i=1}^{k} d(\mu,\psi_i)^2$ is called the Karcher variance of that set. The search for $\mu$ is performed using a gradient approach where an estimate is iteratively updated according to: $\mu \rightarrow \exp_\mu(\epsilon v)$, $v = \frac{1}{k}\sum_{i=1}^{k}\exp_\mu^{-1}(\psi_i)$. Here, exp and $\exp^{-1}$ are as given in the previous section, and $\epsilon > 0$ is a small number. The gradient process is initialized to $\bar{\psi}/\sqrt{\langle\bar{\psi},\bar{\psi}\rangle}$, where $\bar{\psi} = \frac{1}{k}\sum_i \psi_i$.

In Figure 3, we show two examples of Karcher means. Column (a) shows examples of sampling functions $\gamma_1,\gamma_2,\ldots,\gamma_{10}$, and column (b) shows their Karcher means $\mu_\gamma$ (the sampling function obtained by squared integration of $\mu\in\Psi$).

### 3.4. Probability Distributions and Simulations

Having established a representation and a Riemannian metric on the space $\Gamma$ of sampling functions, we now turn to the question of constructing a probability distribution. Recall that a sampling $s$ is a triple $\langle n,\tau,\gamma\rangle \in \mathbb{N}\times\mathbb{S}^1\times\Gamma$. We can write the probability for $s$ as $P(s|C) = P(n)P(\tau|C)P(\gamma|\tau,C)$; we will use a geometric distribution for $n$. The most interesting part of the distribution is the factor $P(\gamma|C,\tau)$. Clearly the possibilities here are enormous. We will restrict ourselves to



(a) (b) (c)

**Fig. 4**. Curvature-driven sampling: (a) a curve; (b) a smoothed version, with $\exp(|\kappa(s)|/\rho)$ displayed as a normal vector field; (c) $\gamma_q$.

"Gaussian" distributions of the form

$$P(\gamma|\tau,C) = Z^{-1}e^{-\frac{1}{2\sigma_s^2}d^2(\dot{\gamma}^{\frac{1}{2}},\psi_0)} \ , \qquad (2)$$

where $d$ is the geodesic distance under our chosen Riemannian metric, and where $\psi_0 = \dot{\gamma}_0^{\frac{1}{2}}$ is, in consequence, the mode of the distribution. We discuss two possibilities for $\gamma_0$ and $\sigma_s$.

The simplest possibility is to emphasize the samplings of a curve that are uniform with respect to its arc-length parameterization, independently of $C$, by choosing $\gamma_0(s) = s$, or equivalently $\psi_0 \equiv 1$. Alternatively, $\gamma_0$ may depend on local geometrical properties, *e.g.* sampling density may increase with increasing curvature of the underlying curve. Define $E(s) = \int_0^s \exp(|\kappa(s')|/\rho)\,ds'$, where $\kappa(s')$ is the curvature of $q$ at arc-length parameter point $s'$ and $\rho\in\mathbb{R}^+$ is a constant. The ratio $\gamma_I(s) = E(s)/E(1)$ is a diffeomorphism, from $[0,1]$ to itself, and the desired sampling for that curve is $\gamma_q = \tau + \gamma_I^{-1}$. The inverse of $\gamma_I$ can be numerically estimated using a spline interpolation. To define a single $\gamma_0$ for each class, we use training curves from that class. First we compute $\gamma_q$ for each training curve, and then, using the techniques presented in Section 3.3, we compute their Karcher mean, which we use as $\gamma_0$, using the Karcher variance as $\sigma_s^2$. We now illustrate these ideas with some examples.

Shown in Figure 4, column (a), are two shapes $q$. We smooth these curves using Gaussian filters: their smoothed versions are shown in column (b). For these smoothed curves, we compute $\kappa$ and then $E(s)$. This function is displayed as a normal vector field on the smoothed curve in (b). Finally, $\gamma_q$ is computed; it is shown in column (c). Figure 5 shows some examples of class-specific means of the $\gamma_q$ for two classes. By using these means as $\gamma_0$ for each class, we can form class-specific priors of the form given in Eqn. 2.

To simulate from probability densities of the form in Eqn. 2, we first randomly generate a function $f\in T_{\psi_0}(\Psi)$ such that $|f| = 1$, where, as before, $\psi_0 = \dot{\gamma}_0^{\frac{1}{2}}$. Then, we generate a normal random variable $x\sim N(0,\sigma_s^2)$, and compute

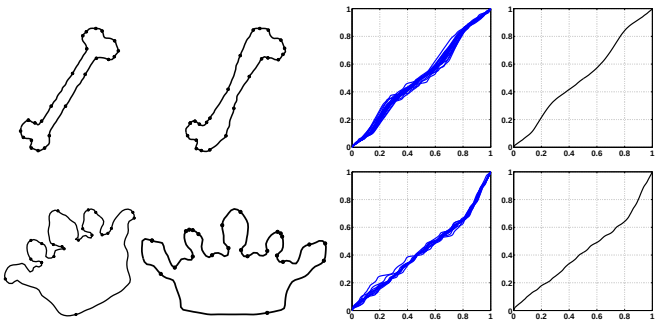


**Fig. 5**. Each row shows two examples of training curves in a class, the sampling functions $\gamma_\kappa$ for that class, and their Karcher means.

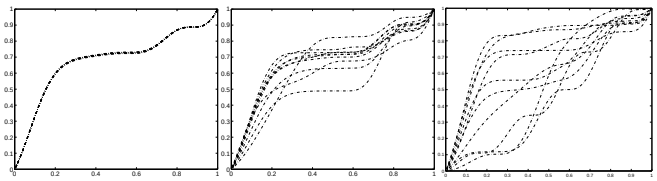


**Fig. 6**. Random samples from $P(\gamma|C)$ with $\sigma_s^2$ increasing from left to right.

a point $\psi = \cos(x)\psi_0 + \sin(x)f/\|f\|$. The random sampling function is then given by $\gamma(s) = \int_0^s \psi(s')^2\, ds'$. Figure 6 shows some examples of random simulations from such a class-specific prior density for increasing values of $\sigma_s^2$.

## 4. SHAPE AND SHAPE VARIABILITY

We now turn to the construction of the shape model, $P(q|C)$. While objects of a given class are similar in their shapes, there is naturally also variability within each class. It is this commonality and variability that $P(q|C)$ must describe. There have been several recent papers that develop tools for analyzing the shapes of planar closed curves, *e.g.* [2, 3]. The main differences amongst these articles lie in the choice of representation for the curves and of the metric used to compare shapes. Two recent papers [7, 8] present an efficient representation under which an elastic metric becomes a simple $\mathbb{L}^2$ metric, with the result that shape analysis simplifies considerably. This has been called the square-root elastic framework, and we describe it briefly here.

Consider a closed, parameterized curve, a differentiable mapping $\beta$ from $\mathbb{S}^1$ to $\mathbb{R}^2$, whose shape we wish to analyze. As described in [7, 8], we will represent a curve $\beta$ by its square-root velocity function: $q : \mathbb{S}^1 \to \mathbb{R}^2$, where $q(t) = \frac{\dot\beta(t)}{|\dot\beta(t)|^{\frac{1}{2}}}$, $|\cdot|$ is the Euclidean norm in $\mathbb{R}^2$, and $t$ is an arbitrary coordinate on $\mathbb{S}^1$. Restricting to unit length, closed curves, we obtain $\mathcal{C} = \{q|\ \int_{\mathbb{S}^1}(q(t)\cdot q(t))dt = 1,\ \int_{\mathbb{S}^1} q(t)\|q(t)\|dt = 1\}$, where $(\cdot)$ is the Euclidean inner product in $\mathbb{R}^2$.

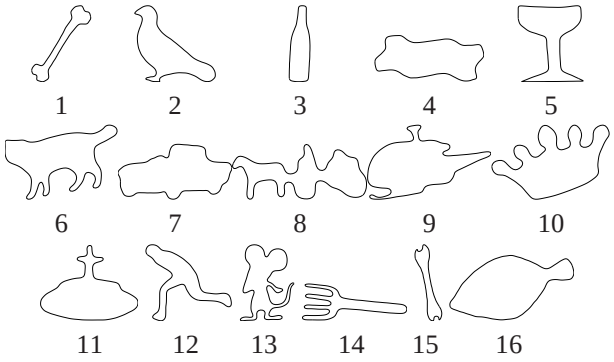



**Fig. 7**. Karcher means of the 16 shape classes used.

We want our shape analysis to be invariant to rigid motions, uniform scaling and re-parameterizations of the curves. The translation and scaling groups have already been removed in defining $\mathcal{C}$. The remaining two: rotations $SO(2)$ and reparameterizations $\mathrm{Diff}(\mathbb{S}^1)$ are removed as follows. Since the actions of these two groups on $\mathcal{C}$ are isometric, with respect to the $\mathbb{L}^2$ metric, we can define the shape space to be the quotient space $\mathcal{Q} = \mathcal{C}/(\mathrm{SO}(2) \times \mathrm{Diff}(\mathbb{S}^1))$ and inherit the $\mathbb{L}^2$ metric from $\mathcal{C}$. In other words, for a point $q \in \mathcal{Q}$ the Riemannian metric takes the form $\langle \delta q_1, \delta q_2 \rangle_q = \int_{\mathbb{S}^1} \delta q_1(t) \cdot \delta q_2(t)\, dt$. To perform statistical analysis in $\mathcal{Q}$, however, which is our goal, one needs to construct geodesics in $\mathcal{Q}$. Joshi *et al*. [8] describe a gradient-based technique for computing geodesics in $\mathcal{Q}$. The technique uses path-straightening flows: a given pair of shapes is first connected by an initial, arbitrary path that is then iteratively "straightened" so as to minimize its length [7]. The length of the resulting path is then the geodesic distance between the shapes. Since one of the effects of $\mathrm{Diff}(\mathbb{S}^1)$ is different placements of the origin on closed curves, its removal results in an alignment of shapes in that regard. One can define and compute the mean of a collection of shapes using the Karcher mean, now based on the geodesic computations [9]. The Karcher means for all the 16 classes used in later experiments are displayed in Figure 7.

The next step is to impose a probability model on $\mathcal{Q}$. Perhaps the simplest model is the one used for $\Gamma$, Eqn. 2. As suggested in [9], it is much easier to express this distribution using the tangent space $T_{q_0}\mathcal{Q}$ to $\mathcal{Q}$ at the mean shape $q_0$ than using $\mathcal{Q}$ itself, because the former is a vector space. In that space, one can use principal component analysis (PCA) and impose a standard Gaussian distribution on the PCA coefficients, then use the exponential map to "push forward" these tangent vectors to $\mathcal{Q}$ itself. Empirical study shows, however, that the histograms of these tangent principal coefficients are often far from Gaussian. We therefore use kernel estimates of the underlying densities to capture this more complex behavior. To simulate from $P(q|C)$ described above, we first

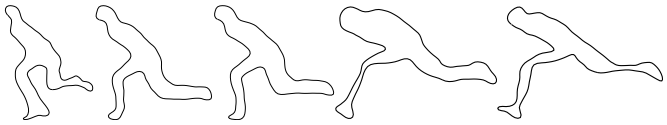


**Fig. 8**. Some random shapes from a TPCA model.

simulate from the estimated density of the tangent principal coefficients, and then use the exponential map to generate the corresponding elements of $\mathcal{Q}$. Figure 8 shows some examples of simulations from one such non-parametric model.

## 5. EXPERIMENTAL RESULTS

Shown in the top three rows of Figure 9 are experimental results on the simulated data with $m = 40$ and $n_0 = 20$. In each case, the left panel shows the true underlying curve which was sampled to generate the data $\mathbf{y}$ which are also shown there. The next panel displays a bar chart of estimated $P(C_i|\mathbf{y})$ for this $\mathbf{y}$, $i = 1, 2, \ldots, 16$ using $J = 300$ samples. The last figure shows a high probability polygon formed using the subsets $\mathbf{y}_s$. In each of the three cases, the amount of clutter is quite high – the number of points on the curve equals the number of clutter points. Still, the algorithm puts the highest probability on the correct class for all cases. As these experiments suggest, the algorithm is able to put high probability on the correct shape class despite the presence of clutter.

## 6. SUMMARY

We have presented a Bayesian approach for finding shape classes in a given configuration of points that is characterized by under sampling of curves, observation noise, and background clutter. Rather than trying all possible permutations of points, we take a synthesis approach and simulate configurations using prior models on shape and sampling. The class posterior is estimated using a Monte Carlo approach.

## 7. REFERENCES


[1] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models: Their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[2] E. Klassen, A. Srivastava, W. Mio, and S. Joshi, "Analysis of planar shapes using geodesic paths on shape spaces," *IEEE Patt. Analysis and Machine Intell.*, vol. 26, no. 3, pp. 372–383, March, 2004.

[3] P. W. Michor and D. Mumford, "Riemannian geometries on spaces of plane curves," *Journal of the European Mathematical Society*, vol. 8, pp. 1–48, 2006.

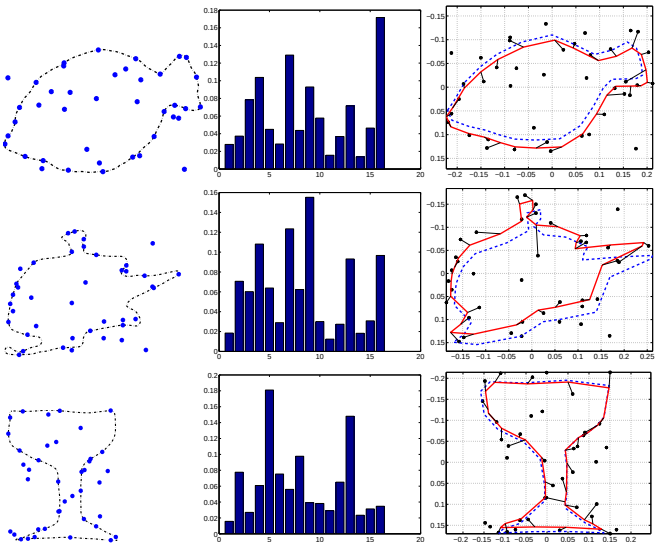



**Fig. 9**. Top three rows – The original curve and the simulated dataset $\mathbf{y}$ (left), and the estimated posterior $P(C_i|\mathbf{y})$ (middle), and a high-probability configuration (last). The correct classes in these examples are: 16, 9, and 5. The bottom left plot shows the average classification performance versus $\nu$ for the Bayesian approach, while the bottom right compares this approach with a classification that uses the Hausdorff metric and an ICP algorithm.


[4] N. N. Čencov, *Statistical Decision Rules and Optimal Inferences*, vol. 53 of *Translations of Mathematical Monographs*, AMS, Providence, USA, 1982.

[5] A. Bhattacharya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bull. Calcutta Math. Soc.*, vol. 35, pp. 99–109, 1943.

[6] S. Amari, *Differential Geometric Methods in Statistics*, Lecture Notes in Statistics, Vol. 28. Springer, 1985.

[7] S. H. Joshi, E. Klassen, A. Srivastava, and I. H. Jermyn, "A novel representation for efficient computation of geodesics between $n$-dimensional curves," in *IEEE CVPR*, 2007.

[8] S. H. Joshi, E. Klassen, A. Srivastava, and I. H. Jermyn, "Removing shape-preserving transformations in square-root elastic (SRE) framework for shape analysis of curves," in *EMMCVPR, LNCS 4679*, A. Yuille et al., Ed., 2007, pp. 387–398.

[9] A. Srivastava, S. Joshi, W. Mio, and X. Liu, "Statistical shape analysis: Clustering, learning and testing," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 590–602, 2005.